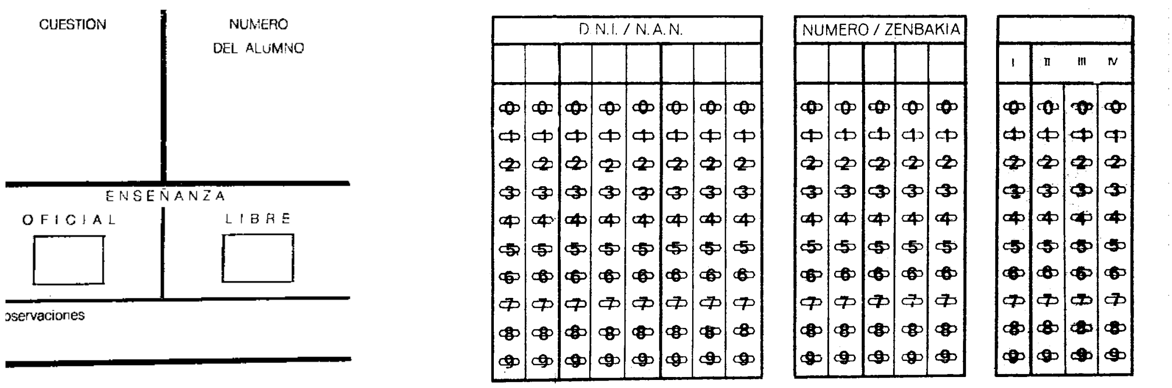
ESTADISTICA EMPRESARIAL - Segundo Curso Curso 2.008-09
Convocatoria de Junio. 08-06-09

## INSTRUCCIONES

1. El examen consta de cuestiones, que se responden sobre la hoja de codificación proporcionada, y problemas, que se responden en papel aparte.
2. Para escoger una respuesta, basta efectuar una marca **rellenando debidamente el rectángulo sobre el que está la letra escogida** en la hoja de codificación. Piénsalo antes; aunque puedes borrar si escribes con lápiz (número 2 o similar), marcas que no estén perfectamente borradas pueden ser leídas. Te aconsejamos que señales sobre el formulario de examen las respuestas que te parezcan adecuadas, y emplees los últimos diez minutos del tiempo asignado en transcribirlas a la hoja de codificación.
3. Hay siempre, en las preguntas de elección múltiple, una **única** respuesta correcta. Todas las cuestiones correctamente resueltas valen 1 punto mientras que las fallidas suponen una penalización de 0.2 puntos. Las preguntas no contestadas no suponen penalización.
4. Cada uno de los problemas, A, B y C debe responderse en una hoja de papel diferente. La recogida se producirá escalonadamente, en los momentos que constarán en la pizarra; primero, la hoja de codificación, y luego los problemas A, B, y C en este orden.
5. El formulario de examen tiene seis hojas numeradas correlativamente al pie (del 0.1 al 0.6). Cerciórate de recibirlas todas, y reclama si tu formulario fuera incompleto. Hay distintos tipos de examen. Este es del tipo 0; marca un 0 en la columna I de tu hoja de codificación, como en el ejemplo.
6. Los puntos obtenibles en cuestiones y problemas son 30 y 30 respectivamente. Son precisos 15 y 15 para superar el examen. Cuestionarios con puntuaciones iguales o superiores a 14 pueden en algún caso ser compensados por una buena nota en los problemas.
7. Rellena tus datos en la hoja de codificación y pliegos de papel suministrados. En "Convocatorias" (columna II) pondrás el número de convocatorias consumidas *incluyendo ésta.*

Ejemplo:

$\underbrace{12545}$ PEREZ, Ernesto Examen tipo $\underbrace{\ 0\ }$ Convocatorias



## CUESTIONES (Duración: 1 hora 30 minutos)

1. **PREGUNTA–REGALO.** La capital de España es:

   (A) París (B) Sebastopol (C) Madrid (D) Londres (E) Pekín

**Las cuestiones 2 a 4 hacen referencia al siguiente enunciado:**

El ancho en milímetros de las componentes de un dispositivo eléctrico en un taller sigue una distribución uniforme $U(5, 15)$. Se consideran aceptables todas aquellas componentes cuyo ancho quede comprendido en el intervalo (8,12).

2. Si en un periodo determinado se han producido 20 componentes, la probabilidad de que al menos 8 de ellas se consideren aceptables es:

   (A) 0.4044 (B) 0.4159 (C) 0.2447 (D) 0.5841 (E) 0.5956

3. En ese periodo de tiempo, el número esperado de componentes aceptables producidas por el taller es:

   (A) 10 (B) 6 (C) 8 (D) 12 (E) 2

4. Si en otro periodo de tiempo se han producido 100 componentes, la probabilidad aproximada de que en este nuevo periodo de tiempo se **rechacen** menos de 51 componentes es:

   (A) 0.0613 (B) 0.9750 (C) 0.0262 (D) 0.9838 (E) 0.025

**Las cuestiones 5 a 7 hacen referencia al siguiente enunciado:**

El número de personas que acude por minuto a una determinada oficina gubernamental sigue una distribución de Poisson con varianza igual a 3. Se asume independencia entre las llegadas en diferentes minutos.

5. La probabilidad de que en un minuto acudan a la oficina gubernamental al menos 3 personas es:

   (A) 0.4232 (B) 0.8153 (C) 0.3528 (D) 0.6472 (E) 0.5768

6. La probabilidad de que en el transcurso de 3 minutos acudan como mucho 7 personas es:

   (A) 0.7440 (B) 0.4557 (C) 0.9881 (D) 0.3239 (E) 0.6761

7. La probabilidad aproximada de que en media hora acudan como mucho 118 personas es:

   (A) 0.014 (B) 0.9750 (C) 0.9987 (D) 0.025 (E) 0.0019

8. Sean $X$ e $Y$ dos v.a. independientes y tales que $X \in \gamma(1/2, 1)$ e $Y \in \gamma(1/2, 1)$. La distribución de la v.a. $X + Y$ es:

   (A) $\chi_2^2$ (B) $\chi_4^2$ (C) $\gamma(1, 2)$ (D) Todo falso (E) $\gamma(1/2, 4)$

9. Sea $\{X_n\}_{n\geq 2}$ una sucesión de variables aleatorias definidas como:

$$X_n = \begin{cases} -n, & \text{con probabilidad } \frac{1}{n} \\ 0, & \text{con probabilidad } 1-\frac{2}{n} \\ n, & \text{con probabilidad } \frac{1}{n} \end{cases}$$

La sucesión converge:

(A) Sólo en probabilidad a $X = 0$

(B) Sólo en distribución a $X = 0$

(C) Sólo en distribución y probabilidad a $X = 0$

(D) En distribución, probabilidad y media cuadrática a $X = 0$

(E) Todo falso

10. Sea $\{X_n\}_{n\in\mathcal{N}}$ una sucesión de variables aleatorias con distribución $N(4, \sigma^2 = 2/n^2)$. La sucesión convergerá:

(A) En distribución, probabilidad y media cuadrática a $X = 4$

(B) Sólo en probabilidad a $X = 4$

(C) Sólo en distribución a $X = 4$

(D) Sólo en distribución y probabilidad a $X = 4$

(E) Todo falso

11. Sean $X$ e $Y$ v.a. independientes, ambas con distribución $N(0,1)$. Sea $Z = \frac{X}{\sqrt{Y^2}}$. La distribución de $Z$ es:

(A) $t_1$ (B) $N(0,1)$ (C) $\chi^2_1$ (D) $N(0,2)$ (E) $Z \equiv 1$

12. Sea $X$ una variable aleatoria con distribución $t_n$, $t$ de Student con $n$ grados de libertad. Entonces se verifica que:

(A) $t_{n,\alpha} > t_{n,\frac{\alpha}{4}}$ (B) $t_{n,\alpha} < t_{n,\frac{\alpha}{2}}$ (C) $\mathrm{E}(X) = n$ (D) $t_{n,\frac{\alpha}{2}} > t_{n,\frac{\alpha}{4}}$ (E) $t_{n,\alpha} = t_{n,1-\alpha}$

13. Sea $X$ una variable aleatoria con distribución $\mathcal{F}_{10,20}$. Entonces el valor de $k$ tal que $P(X > k) = 0.90$ es:

(A) 0.45 (B) 0.52 (C) 2.20 (D) 0.10 (E) 1.94

**Las cuestiones 14 y 15 hacen referencia al siguiente enunciado:**

Sea $X$ una v.a. con función de cuantía dada por:

$$P(X=0) = \theta^2; \quad P(X=1) = 2\theta(1-\theta); \quad P(X=2) = (1-\theta)^2.$$

Para estimar el parámetro $\theta$ se ha tomado una m.a.s. de tamaño $n = 100$, en la que se ha obtenido 22 veces el cero, 53 veces el uno y 25 veces el dos.

14. La estimación de $\theta$ por el método de máxima verosimilitud es:

(A) 0.485 (B) 0.125 (C) 0.776 (D) 0.225 (E) 0.554

15. La estimación de $\theta$ por el método de momentos es:

(A) 0.554 (B) 0.125 (C) 0.485 (D) 0.225 (E) 0.776

**Las cuestiones 16 a 19 hacen referencia al siguiente enunciado:**

Sea una variable aleatoria con distribución uniforme $U[\theta, \theta+1]$ de la que, para estimar el parámetro $\theta$, se toma una m.a.s. de tamaño $n$, $X_1, \ldots, X_n$. Se propone el estimador $\hat{\theta} = \overline{X} - \frac{1}{2}$. Además, se sabe que la varianza de una v.a. $U[a,b]$ es $\sigma^2 = (b-a)^2/12$.

16. ¿Es este estimador insesgado?

(A) Sí (B) - (C) - (D) - (E) No

17. La varianza de este estimador es:

(A) $\frac{1}{12n}$ (B) $\frac{1}{12n} + 1$ (C) $\frac{1}{12}$ (D) $\frac{1}{12n} + \frac{1}{2}$ (E) Todo falso

18. El error cuadrático medio de este estimador es:

(A) $\frac{1}{12}$ (B) $\frac{1}{12n} + 1$ (C) Todo falso (D) $\frac{1}{12n} + \frac{1}{2}$ (E) $\frac{1}{12n}$

19. ¿Es este estimador consistente?

(A) No (B) - (C) Sí (D) - (E) -

**Las cuestiones 20 a 22 hacen referencia al siguiente enunciado:**

Sea $X$ una v.a. con función de densidad

$$f(x; \theta) = \theta x^{\theta-1}, \quad 0 < x \leq 1, \ \theta > 0$$

Se quiere contrastar la hipótesis nula $H_0 : \theta = 1$ frente a la hipótesis alternativa $H_1 : \theta = 2$. Para ello se ha tomado de dicha población una muestra de un sólo elemento (es decir, se observa $X$).

20. La región crítica de máxima potencia para $X$ será de la forma:

(A) $(K, 1)$ (B) $(K_1, K_2)$ (C) Todo falso (D) $(K_1, K_2)^c$ (E) $(0, K)$

21. Si $\alpha = 0.05$, la región crítica de mayor potencia será:

(A) $(0, 0.05)$ (B) $(0.05, 0.95)$ (C) $(0.95, 1)$ (D) $[0.05, 0.95]^c$ (E) $(0.90, 1)$

22. La potencia del contraste es:

(A) 0.9025 (B) 0.5487 (C) 0.0975 (D) 0.4513 (E) 0.0025

**Las cuestiones 23 a 26 hacen referencia al siguiente enunciado:**

El número de clientes por minuto que llega a una sucursal bancaria es una v.a. con distribución de Poisson de parámetro $\lambda$. Para contrastar $H_0 : \lambda = 1$ frente a $H_1 : \lambda = 2$, se toma una m.a.s. de 4 minutos en los que se observan 3, 2, 1 y 5 clientes respectivamente.

23. Si llamamos $Z$ al número total de clientes que han llegado a la sucursal durante los 4 minutos, la región crítica de mayor potencia será de la forma:

(A) $Z \leq K$ (B) $Z \in [K_1, K_2]$ (C) $Z \geq K$ (D) $Z \in [K_1, K_2]^c$ (E) Todo falso

24. Si $\alpha = 0.05$, la región crítica de mayor potencia será:

(A) $Z \geq 9$ (B) $Z \in [3, 9]$ (C) Todo falso (D) $Z \in [3, 9]^{c}$ (E) $Z \leq 3$

25. La potencia del contraste es:

(A) 0.5925 (B) 0.7166 (C) 0.0424 (D) 0.2834 (E) 0.4075

26. La decisión del contraste será:

(A) Rechazar la hipótesis nula (B) - (C) - (D) - (E) No rechazar la hipótesis nula

**Las cuestiones 27 y 28 hacen referencia al siguiente enunciado:**

En un proceso de fabricación se desea contrastar la hipótesis nula de que el porcentaje de piezas defectuosas fabricadas es no mayor del 10%, frente a la alternativa de que es mayor de 10%. Se toma una m.a.s. de tamaño $n = 200$, en donde se observaron 25 piezas defectuosas.

27. Con una confianza aproximada del 90% se puede afirmar que el porcentaje de piezas defectuosas fabricadas se encuentra en el intervalo:

(A) $(0.05,\ 0.12)$ (B) $(0.11,\ 0.18)$ (C) $(0.15,\ 0.22)$

(D) $(0,\ 0.05)$ (E) $(0.09,\ 0.16)$

28. Al nivel de significación aproximado del 10%, la decisión del contraste será:

(A) Rechazar la hipótesis nula (B) - (C) -

(D) - (E) No rechazar la hipótesis nula

**Las cuestiones 29 y 30 hacen referencia al siguiente enunciado:**

Se quiere estimar el gasto medio en lotería de navidad por familia, expresado en euros, en una determinada provincia. Para ello se ha tomado una m.a.s. de 61 familias de dicha provincia a las que se ha preguntado cuál ha sido el dinero gastado en dicho concepto. A partir de dicha muestra, se ha obtenido que $\overline{x} = 188$ y $s^2 = 10000$. Se asume normalidad.

29. El intervalo del 95% de confianza para el gasto medio en lotería es:

(A) $(166.44,\ 209.56)$ (B) $(154.33,\ 204.56)$ (C) $(162.18,\ 213.82)$

(D) $(178.56,\ 198.65)$ (E) $(180.33,\ 235.77)$

30. Se quiere contrastar la hipótesis nula de que el gasto medio ha sido 200 euros frente a la hipótesis alternativa de que ha sido distinto de dicha cantidad. La decisión al nivel de significación 5% será:

(A) Rechazar la hipótesis nula (B) No se puede decidir (C) -

(D) - (E) No rechazar la hipótesis nula

**PROBLEMAS (Duración: 75 minutos )**

**A.** (10 puntos, 25 minutos)

Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño $n$ tomada de una v.a. $X$ cuya función de densidad de probabilidad viene dada dada por:

$$f(x;\theta) = (\theta + 4)\ \ x^{\theta+3}, \quad 0 < x < 1,\ \theta > 0$$

i) Obtén el estimador para el parámetro $\theta$ por el método de máxima verosimilitud.

ii) Obtén el estimador para el parámetro $\theta$ por el método de momentos.

**B.** (10 puntos, 25 minutos)

Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño $n = 15$ tomada de una población binaria $b(p)$. Queremos contrastar la hipótesis nula $H_0 : p = 0.50$ frente a la hipótesis alternativa $H_1 : p = 0.30$.

i) Deduce la forma de la región crítica de mayor potencia para dicho contraste.

ii) Al nivel de significación del 5%, calcula la región crítica del contraste.

iii) ¿Cuál es la potencia del contraste?

**C.** (10 puntos, 25 minutos) Sea $X$ una v.a. con distribución $N(m, \sigma^2)$ con varianza desconocida, que se quiere estimar. Para ello se ha tomado una m.a.s. de tamaño $n$, $X_1, \ldots, X_n$.

i) Deducir de forma detallada el intervalo de confianza $(1 - \alpha)$ para la varianza poblacional $\sigma^2$.

ii) En el caso de que $n = 30$ y que de la muestra se obtenga el valor de la varianza muestral $s^2 = 25$, obtener el intervalo de confianza 95% para la **desviación típica poblacional** $\sigma$.

**SOLUCIONES DEL CUESTIONARIO (tipo 0)**

| | | |
|---|---|---|
| 1: C | 11: A | 21: C |
| 2: D | 12: B | 22: C |
| 3: C | 13: A | 23: C |
| 4: C | 14: A | 24: A |
| 5: E | 15: C | 25: E |
| 6: D | 16: A | 26: A |
| 7: C | 17: A | 27: E |
| 8: B | 18: E | 28: E |
| 9: C | 19: C | 29: C |
| 10: A | 20: A | 30: E |

## SOLUCIONES DE LOS PROBLEMAS

### Problema A

Sabemos que $X_i,\ i = 1, \ldots, n$, tiene función de densidad de probabilidad dada por:

$$f(x_i;\theta) = (\theta + 4)\ \ x_i^{\theta+3},\quad 0 < x_i < 1,\ \theta > 0$$

**i**) Así, la función de verosimilitud estará dada por

$$L(\vec{x};\theta) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta)$$

$$L(\vec{x};\theta) = \left[(\theta + 4)\ \ x_1^{\theta+3}\right] \left[(\theta + 4)\ \ x_2^{\theta+3}\right] \cdots \left[(\theta + 4)\ \ x_n^{\theta+3}\right]$$

$$L(\vec{x};\theta) = (\theta + 4)^n \left(\prod_{i=1}^{n} x_i\right)^{(\theta+3)}$$

El estimador máximo verosímil de $\theta$ es el valor que maximiza la función de verosimilitud o, equivalentemente, su logaritmo neperiano:

$$\ln L(\vec{x};\theta) = n \ln(\theta + 4) + (\theta + 3) \ln\left(\Pi_{i=1}^{n} x_i\right)$$

Derivando respecto de $\theta$, tendremos que:

$$\frac{\partial \ln L(\vec{x};\theta)}{\partial \theta} = \frac{n}{(\theta + 4)} + \ln\left(\prod_{i=1}^{n} x_i\right) = 0$$

Por lo tanto,

$$\frac{n}{(\theta + 4)} = -\ln\left(\prod_{i=1}^{n} x_i\right),$$

de donde

$$\hat{\theta}_{\mathrm{MV}} = -\frac{n}{\ln\left(\prod_{i=1}^{n} x_i\right)} - 4$$

**ii**) Para calcular el estimador de $\theta$ por el método de momentos, debemos encontrar $\alpha_1 = \mathrm{E}(X)$. En este caso,

$$\alpha_1 = \mathrm{E}(X) = \int_0^1 x f(x;\theta) dx = \int_0^1 x\left[(\theta + 4) x^{\theta+3}\right] dx = \int_0^1 (\theta + 4) x^{\theta+4} dx$$

$$\Longrightarrow \alpha_1 = \mathrm{E}(X) = (\theta + 4) \left[\left.\frac{x^{\theta+5}}{(\theta + 5)}\right|_0^1\right] = \left(\frac{\theta + 4}{\theta + 5}\right)$$

Para calcular el estimador por el método de momentos igualamos los momentos poblacionales y muestrales. Es decir, $\alpha_1 = a_1$. De esta forma,

$$\alpha_1 = \mathrm{E}(X) = \left(\frac{\theta+4}{\theta+5}\right) = \overline{X} = a_1,$$

de donde:

$$\theta + 4 = \overline{X}(\theta+5) \Longrightarrow \theta(1-\overline{X}) = 5\overline{X} - 4 \Longrightarrow \hat{\theta}_{MM} = \left(\frac{5\overline{X}-4}{1-\overline{X}}\right)$$

**Problema B**

Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño $n = 15$ tomada de una población binaria $b(p)$. Queremos contrastar la hipótesis nula $H_0 : p = 0.50$ frente a la hipótesis alternativa $H_1 : p = 0.30$.

**i)** Para deducir la forma de región crítica de mayor potencia hacemos uso del Teorema de Neyman-Pearson. Así, las funciones de verosimilitud bajo las hipótesis nula y alternativa, estarán dadas por:

$$L(\vec{x}; p_0) = L(\vec{x}; p = 0.50) = (0.50)^{\Sigma_{i=1}^n x_i}(1-0.50)^{n-\Sigma_{i=1}^n x_i},$$

y

$$L(\vec{x}; p_1) = L(\vec{x}; p = 0.30) = (0.30)^{\Sigma_{i=1}^n x_i}(1-0.30)^{n-\Sigma_{i=1}^n x_i},$$

respectivamente. Por tanto, si aplicamos el Teorema de Neyman-Pearson, tendremos que:

$$\frac{L(\vec{x}; p_o)}{L(\vec{x}; p_1)} = \frac{(0.50)^{\Sigma_{i=1}^n x_i}(1-0.50)^{n-\Sigma_{i=1}^n x_i}}{(0.30)^{\Sigma_{i=1}^n x_i}(1-0.30)^{n-\Sigma_{i=1}^n x_i}} \leq K,\ K > 0$$

$$\Longrightarrow \left[\frac{(0.50)(1-0.30)}{(0.30)(1-0.50)}\right]^{\Sigma_{i=1}^n x_i} \quad \left[\frac{(1-0.50)}{(1-0.30)}\right]^n \leq K$$

$$\Longrightarrow \left[\frac{(0.50)(1-0.30)}{(0.30)(1-0.50)}\right]^{\Sigma_{i=1}^n x_i} \leq K_1,\ K_1 > 0$$

Tomando logaritmos neperianos, tenemos que:

$$\left(\sum_{i=1}^n x_i\right) \ln\left[\frac{(0.50)(1-0.30)}{(0.30)(1-0.50)}\right] \leq K_2,\ K_2 > 0$$

Ahora, como $0.50 > 0.30$ y, además, $(1-0.30) > (1-0.50)$, el logaritmo es positivo, por lo que tendremos que la regla de decisión será rechazar la hipótesis nula si $\sum_{i=1}^n X_i \leq C$. Por tanto, la forma de la región crítica de mayor potencia para el estadístico de contraste $Z = \sum_{i=1}^n X_i$ será $\mathrm{RC} = [0, C]$.

**ii)** Al nivel de significación $\alpha = 0.05$, teniendo en cuenta que $Z = \sum_{i=1}^n X_i \in b(p, 15)$, tendremos que:

$$\alpha = 0.05 \geq P[Z \in \mathrm{RC}|H_0] = P[Z \leq C | Z \in b(0.50, 15)] = F_Z(C)$$

$$\Longrightarrow F_Z(C) \leq 0.05 \Longrightarrow C = 3 \Longrightarrow \mathrm{RC} = [0, 3].$$

Es decir, rechazamos la hipótesis nula si $Z = \sum_{i=1}^n X_i \leq 3$.

**iii)** Para calcular la potencia tendremos que:

$$\text{Pot} = P[Z \in \text{RC}|H_1] = P[Z \leq 3|Z \in b(0.30, 15)] = F_Z(3) = 0.2969.$$

**Problema C**

**i)** Sea $\vec{X} = (X_1, X_2, \ldots, X_n)$ una m.a.s. de tamaño $n$ tomada de una distribución $N(m, \sigma^2)$. Entonces:

$$\frac{nS^2}{\sigma^2} \in \chi^2_{n-1}$$

**Intervalo de Confianza**

$$P\left(\chi^2_{n-1|1-\alpha/2} < \frac{nS^2}{\sigma^2} < \chi^2_{n-1|\alpha/2}\right) = 1 - \alpha$$

$$P\left(\frac{1}{\chi^2_{n-1|\alpha/2}} < \frac{\sigma^2}{nS^2} < \frac{1}{\chi^2_{n-1|1-\alpha/2}}\right) = 1 - \alpha$$

$$P\left(\frac{nS^2}{\chi^2_{n-1|\alpha/2}} < \sigma^2 < \frac{nS^2}{\chi^2_{n-1|1-\alpha/2}}\right) = 1 - \alpha$$

Por lo tanto, el intervalo de confianza $1 - \alpha$ para la varianza es:

$$IC_{1-\alpha} = \left(\frac{ns^2}{\chi^2_{n-1|\alpha/2}}, \frac{ns^2}{\chi^2_{n-1|1-\alpha/2}}\right)$$

**ii)** En el caso de que $n = 30$ y $s^2 = 25$ el intervalo de confianza 95% para la varianza será:

$$IC_{1-\alpha} = \left(\frac{750}{\chi^2_{n-1|\alpha/2}}, \frac{750}{\chi^2_{n-1|1-\alpha/2}}\right)$$

Dado que $\chi^2_{29|0.05/2} = 45.7$ y $\chi^2_{29|1-0.05/2} = 16$, entonces, el intervalo de confianza 95% para la varianza será:

$$IC_{0.95} = (16.41, 46, 88),$$

de donde, el intervalo de confianza 95% para la desviación típica será:

$$IC_{0.95} = (\sqrt{16.41}, \sqrt{46, 88}) = (4.05, 6.85).$$